A tale of two tails:

Do Power Law and Lognormal models fit firm-size distributions in the mid-Victorian era?

Piero Montebruno[†][*], Robert J. Bennett [‡][*], Carry van Lieshout[§][*], and Harry Smith[¶][*]

[*]University of Cambridge, Department of Geography and Cambridge Group for the History of Population and Social Structure, Downing Place, Cambridge, CB2 3EN, UK.

February 25, 2019

[†]pfm27@cam.ac.uk
[‡]Corresponding author. rjb7@cam.ac.uk
[§]cv313@cam.ac.uk
[¶]hjs57@cam.ac.uk

Abstract

The paper explores the frequency and size distributions of firm-size in a novel dataset for the mid-Victorian era from a recent extraction of the England and Wales population censuses of 1851, 1861, 1871, and 1881. The paper contrasts the hypothesis of the Power Laws against the Lognormal model for the tails of the distributions using maximum likelihood estimation, log likelihood ratio, clipped sample coefficient of variation UMPU-Wilks test, Kolmogorov-Smirnov statistic, among other state-of-the-art statistical methods. Our results show that the Power Law hypothesis is accepted for the size distribution for the years 1851 and 1861, while 1871 is marginally non-significant, but for 1881 the test is inconclusive. The paper discusses the process that generates these distributions citing recent literature that shows how after adding an i.i.d. noise to the Gibrat's multiplicative model one can recreate a Power Law behaviour. Overall, the paper provides, describes and statistically tests for the very first time a unique historical dataset confirming that the tails of the distributions at least for 1851 and 1861 follow a Pareto model and that the Lognormal model is firmly rejected.

1

A tale of two tails:

Do Power Law and Lognormal models fit firm-size distributions in the mid-Victorian era?

1   Introduction

Fujita et al. in their classic textbook observe that often "theory gives simple, sharp-edged predictions, whereas the real world throws up complicated and messy outcomes".[17] However, Fujita et al. also notice that numerous examples occur where the situation is upside down: "data offer a stunningly neat picture, one that is hard to reproduce in any plausible (or even implausible) theoretical model".[17] This unsettled theorisation may explain that a search in Physica A gets 7,543 results for "Power law", 551 for "Pareto distribution", 222 for "Zipf's Law", 708 for "Lognormal" and 70 for "Gibrat's Law", the main fields discussed in this paper. In this literature on firm size, the key sources are the work of American linguist George Kingsley Zipf [63] and French engineer Robert Gibrat [25]. However, Power law, Pareto distribution, Zipf's law, and Gibrat's law have been described in an astonishing number of occasions in the social sciences, economics, biology, linguistics, history, geography, ecology, physics, and elsewhere.[1]

This paper focuses on firm size. This has been previously explored using different variables according to the frequency, number or percentage of total employees, employment share, sales, income, assets or profits. The unique data examined in this paper permits analysis of both the distributions of the frequency, and the size, measured by number of total employees of firms in England and Wales between 1851 and 1881. The relevant literature on the distribution of firm size covers almost all geographical contexts: e.g. [51] for the United Kingdom, [5] and [57] for the United States, [22] and [21] for the G7 countries, [46] for Japan, [10] for Italy, [47] for Portugal, [31] and [62] for China, or [33] for Korea. A related topic, farm size, which is also available in our dataset but is not analysed in this paper, is included in works by [27] and [1] for England and Wales, [28] for South Africa, and

---

[1]Mitzenmacher [44] gives nineteen examples in computer science and apologizes for "leaving out countless further examples". Clauset et al [13] give twenty-four examples comprising words in Moby Dick [45], protein interaction in Saccharomyces cerevisiae [35], metabolites in Escherichia Coli [34], intensity of wars [50], terrorist attacks [12], the size distribution of web files transmitted over the Internet [61], species per genus of mammals [56], customers affected in electrical blackouts [45], bestselling books sold [29], acres of wildfires [45], solar flares [45], and number of citations listed in the Science Citation Index [49].

[42] for England and the EU. An important contribution to the discussion is the book [53] where complex stochastic processes are used to explain both Zipf's and the Gibrat's laws. Advanced econometric techniques are implemented in this paper to test the tail behaviour of our unique sample. The main intention of the paper is to provide evidence of the statistical behaviour of the firm size and frequency distributions using a dataset for England and Wales in mid-Victorian times.

Currently there is a heated debate about the possibility to distinguish whether behaviour at the tail follows a Power Law distribution or a pure Gibrat's Law, i.e. a Lognormal distribution. Malevergne et al. [43] claim that their attempt using the UMPU (uniformly most powerful unbiased)—following Del Castillo and Puig test [15]—"is shown to provide a clear diagnostic, allowing us to distinguish between the power-law and the Lognormal hypothesis, even when the data set is quite small". Also, although Clauset et al. [13] argue that: "[i]t is extremely difficult to tell the difference between log-normal and power-law behaviour ... [and] it appears unlikely that any test would be able to tell them apart unless we had an extremely large data set", notwithstanding they provide an extended methodology to deal with the testing of the Power Law and Lognormal hypotheses. Bee et al. [8] and [7] also acknowledge that "the two different distributions are mathematically different, but only in the limit" and the tail behaviour for their empirical analysis gave approximately the same outcomes for their finite samples but at the same time they provide a Maximum Entropy (ME) test to tackle this query. In this paper, we implement several advanced statistical methods proposed in recent literature using maximum likelihood estimation to estimate the shape (α) and the cut-off (u) parameters, loglikelihood ratio to compare competing models, Vuong methodology to find p-values for this comparison, clipped sample coefficient of variation UMPU-Wilks test to contrast the Pareto to the Lognormal hypothesis on the tail, Kolmogorov-Smirnov procedure to measure the distance of models and find also the u parameter and the Goodness-of-Fit, and non-parametric bootstraping to obtain confidence intervals as proposed by Malevergne et al. [43], Clauset et al. [13], Alstott et al. [2], and Gillespie [26]. The evidence we provide corroborates the claim of Malevergne et al. [43] that "the test can be derived with extremely high accuracy (even for very small samples)".

We are interested in which of either a Power Law or Lognormal better describes the distributions of our data, but to aid interpretation it is important to clarify why firm size

follows these distributions? Traditionally for examination of the whole distribution focus was given to the simple distinction of whether a firm's decision-making is multiplicative and not additive. When a decision is made on the next period of workforce hiring, it is based on "I need ten more or twenty fewer workers", or it is based on "I need an increase of 10%, or I need a 20% decrease". The former was considered to generate a normal distribution, while the latter was believed to generate a Lognormal distribution. As a result, many authors explain the outcomes of the Lognormal simply as a proportional process that acts recursively over time. For example, Gan et al. [23] using Monte Carlo experiments showed that no explanatory economic theory is required. Batty [6] argues that this result from complex systems is self-organising and Corominas-Murtra and Sole [14] establish that Zipf is a common statistical feature when a system goes from order to disorder or vice versa. Arshad et al. [3] give a wider comparison of these simple explanations based on economic shocks, human capital accumulation, or central place theory. At the same time, Gabaix [19] and [20] presents an explanation for the growth model of cities. Nevertheless, more recently Saichev et al. [53] and [52] have shown that simply adding a well-behaved noise to the well-known equation embodying Gibrat's law can result in switching the Lognormal behaviour to a Power Law for large values in the tail. That is, there is no need to add a theory just to explain the generation of the Lognormal distribution, as commonly argued, but also for the Power Law.

2   Methodology, data, frequency and size distributions

The data for this paper are derived from the original manuscripts of the England and Wales population census. The censuses have been transcribed and encoded into a database recently made available at the UKDA [32]: The Integrated Census Microdata (I-CeM). From this source as well as additional census data to infill gaps in I-CeM, employers and the size of their workforce have been extracted and made available as the 'British Business Census of Entrepreneurs' database.[2] This provides the starting point here. The 1851-81 population censuses questions included data on workforce size. Unfortunately, the question was discontinued after 1881 so our analysis is limited to four-censuses: 1851-81. The data are particularly valuable since they are exhaustive, especially for the largest firms: noted as

---

[2]Reference to be included when database is deposited in March-April 2019. The data used are included with this submission as supplementary files.

essential by Fujiwara et al. [18] to discriminate between firm size distributions. However, whilst the data for 1851, 1861 and 1881 are almost fully complete, 1871 has transcriber deficiencies that mean that total business numbers are not reliably available (about 20% of data are missing). Hence, whilst 1871 analysis is included here for completeness, and it generally confirm the same distribution patterns, we do not always attempt to fully interpret this year.

There are two possible distributions that can be built from the employers' answers in the censuses. One is the frequency distribution of firm size, the number of firms for a given workforce size. For instance, in 1851 there were 181,900 firms and 390 different sizes. Their distribution starting from the left (largest) has 159 (40.77%) of frequency one—159 sizes appear just once—52 (13.33%) of frequency two—52 sizes appear twice—and 23 (5.9%) of frequency three—23 sizes appear thrice—and so on, ending to the right with one (0.26%) of frequency 38,111 (i.e. for size of two employees), and one (0.26%) of frequency 47,231 (for size of one employee). This is similar to the number of journal citations used in the data by Redner (1998), as well as many other examples. The second distribution is the firm size distribution, the workforce size for a given firm. For instance, in 1851, there are 1,185,602 employees and 181,900 firms and its distribution starting from the left has 47,231 (25.97%) with one employee, 38,111 (20.95%) with two employees, and so on, ending to the right with one with 4,287 employees and one with 6,000 employees. As a starting point, it is possible to plot both together. Figure 1, which is a replication of Figure 1 of Redner [49] for our firm data, using a log-log scale to represent the relation between frequency and size. These plots represent frequency as a function of size. However, the function is not one-to-one because for some low frequencies there are many sizes with the same given frequency, as can be seen from the large cluster of data points at low frequencies and relatively large sizes. This is amplified in our data because of bunching: contrary to Redner's data, employers often reported their workforces in round numbers of tens, hundreds and thousands. This explains the higher variance and spread in frequency at large firm sizes. Note also that we use the actual firm size for each firm: they are not pre-grouped into size classes. This overcomes one of the criticisms of many analyses of firm size: what Bee et al. [8] term "the widespread practice of binning the data" (grouping into size categories). This loses information and results in any statistical inferences being less reliable, especially for the tails which Fujiwara [18] observe are critical: where bins are large and number of observations is small over a wide size range.

4

Tables 1 and 2 show the most frequent and the largest firm sizes for 1851-1881. First notice that, in grey, the function converting rank to size (and its inverse) is not necessarily monotonic. Also, in 1851 the ratio of each frequency relative to the frequency of the most frequent firm size ranges from one to fifteen, but in the rest of the years it ranges approximately one to eight. If we plot the frequency and the size distribution for values of less than forty for each distribution, in Figure 2, more than 40% of the distribution is made up by firms with one specific workforce category, with a large drop of frequency for size classes with two specific size classes and over. Also, in Figure 3, over 20% of the firms have size one, but the drop for smaller sizes is smooth and not as sharp as for frequencies. However, for both distributions what matters most for the mathematics of the curve is the upper tail of the distribution where high figures appear. Figure 4 shows the whole distribution of firm size frequency. For the upper tail there are only 159 of firms including the four largest firms (2,010, 2,180, 4,287, and 6,000 employees). Hence, the distribution does not have a typical scale or size but spans from frequency one to frequency almost 50,000 as described by Newman [45]: "not all things we measure are peaked around a typical value. Some vary over an enormous dynamic range, sometimes many orders of magnitude." In the US population, Newman calculates that the ratio of largest and smallest population is at least 150,000. This compares with our almost 50,000 between the most frequent and least frequent firm size for the approximately 500 different frequency categories of firms. For the size distribution, this compares with approximately 8,000 as the size range spans from one to 8,000 employees if we expand the analysis to all the distributions.

Table 3 compares in detail the summary statistics of the two distributions in each of the four censuses. While the frequency distribution has only around 500 datapoints, the size one has about 180,000. As a result, in the frequency distribution the arithmetic mean is of a different order of magnitude than the geometric and harmonic means and the median. In contrast, in the size distribution all are similar and close together. All statistics for maximum, range, inter quantile range, standard deviation (and variance) and standard error of the mean of the frequency distribution are greater than for firm size. While the size distribution has a greater skewness and kurtosis, both distributions exhibit similar coefficients of variation.

Table 4 shows the correlations among the frequency and size distributions for each of the four censuses. At first sight, there is a high correlation within the distributions: the four frequency distributions are correlated at about 0.99, while the four size distributions are correlated at 0.95. The correlations across types are also high at 0.7 and above.

3   Empirical analysis

A relationship $y = f(x)$ follows a Power-law if, after the unit increase of x, y increases by a power. This Power Law is an intermediate rate of increase between the linear and the exponential. If you take the logs of the dependent variable and also the logs of the independent variable, there is a straight line relationship. The exponential, a model with a higher rate of growth just requires the logs of the dependent variable to get a straight line. Thus, finding the linear behaviour *after taking logs* was one of the holy grails for identifying the type of growth, but as we will see this does not suffice as a test. Indeed, a straight line is only a necessary but not a sufficient behaviour of a Power Law. In our cases, the Lognormal presents an exponential decline while the Power Law just a power decay. Consequently, the Lognormal goes to zero faster than the Power Law. Nevertheless, it is also well-known that for certain parameter values, the Lognormal can present pseudo-linear patterns in the meaningful intervals of analysis. As shown by Malevergne et al. [43] a Lognormal with standard deviations in the interval [2,3] "is close to linear over almost four decades" of city sizes over time. This is observed in our data. Thus it is of crucial importance to use other more advanced methods to distinguish between the Power Law and the Lognormal in the dataset.

Consequently, we have two competing behaviours: Pareto (a Power Law with a cut-off—the x_{min} or u parameter—defining the tail as shown below) and Lognormal distributions.[3] The probability density function of the Pareto distribution as presented in equation

---

[3]It may be confusing the jump between Pareto Law, Pareto, and Zipf. To clarify this, we argue, as will become clear below, that Power Law includes Pareto and that Pareto includes Zipf. This means that when we are saying Pareto we are also saying "Power Law with a cut-off—the x_{min} or u parameter—". So these terms are both interchangeable, but also subtlety different. The title of the paper uses Power Law because it is the more general terminology but the tail behaviour is just restricted to Pareto because we are thinking of a "Power Law with a Lentement term equals to x_{min}". At the same time, when we explore the Zipf behaviour we are thinking of a "Pareto with an exponent equal to 1" or a "Power Law with a cut-off and an exponent equal to 1".

2.2 of Clauset et al. [13] is:

$$p(x) = \frac{\alpha - 1}{x_{min}} \left(\frac{x}{x_{min}} \right)^{-\alpha}$$

while Malevergne et al. [43] equation (21) use:

$$\alpha_0 \frac{u^{\alpha_0}}{x^{1+\alpha_0}}, x > u \tag{3.1}$$

It is immediate that:

$$\alpha = 1 + \alpha_0$$

and that

$$x_{min} = u$$

Moreover, the Power-law, Pareto, and Zipf are related by the following relation

$$Power\text{-}law \supset Pareto \supset Zipf$$

By definition, a Power Law has a complementary cumulative distribution function (CCDF) equal to $P(X > x) \sim L(x)x^{-\alpha_0}$, where $\alpha_0 > 0$ and $L(x)$ is a slowly varying function—"Lentement" in French—as suggested by Karamata's classical definition [37], [38], and [54]:

$$\lim_{x \to \infty} \frac{L(rx)}{L(x)} = 1$$

From this general equation, one can impose an additional constraint that $L(x) = x_{min}$ so that the Pareto distribution is defined. Finally, if the parameter α_0 in Equation 3.1 is equal to 1 a Zipf distribution is defined. Notice, importantly, that the coefficient in the CCDF is α_0, as defined in Equation 3.1, so that Zipf's law is fulfilled when Malevergne et al. $\alpha_0 = 1$ but by Clauset et al. when $\alpha = 2$. Henceforth, we call the cut-off parameter u as in Malevergne et al. and we use Clauset et al. α so the Zipf's law will be attained for a value of $\alpha = 2$.

Sometimes this relation is firstly presented, as we do in Figures 5 and 7, as a linear

7

relation between the rank, r, and firm size, x, in a log-log plot. But it is important to admit that a linear relation is only a necessary condition for a Power Law behaviour but not a sufficient condition, as mentioned before. Hence, a linear regression does not corroborate a Power Law because Lognormal, exponential (see [13]) and other distributions can manifest themselves as linear in a rank/size log-log plot. The following equation, Equation (2) of Hoon et al. [33], further describes this relation:

$$xr^\beta = A$$

or taking logs,

$$\ln(x) = \ln(A) - \beta \ln(r)$$

In this case Malevergne et al. α_0 corresponds to $1/\beta$ of Hoon et al. So again, Zipf's law is guaranteed with a Hoon et al. $\beta = 1$. These authors add "The larger the value of β [the smaller the value of the Malevergne et al. α_0], the greater the relative size of a large firm (high ranked firm) as compared with a smaller firm (low ranked firm)". Hoon et al. also provide intuition with the following differential equation:

$$\frac{dx}{x} = -\beta \frac{dr}{r}$$

which can be interpreted as "for any two companies in the ranking, the difference in their sizes is proportional to their difference in ranks and β is the proportionality constant". In other words, that the elasticity of the relation between rank and firm size is constant.

The Lognormal model has two parameter μ and σ. According to Johnson et al. [36], x is Lognormally distributed with parameters μ and σ, if the logarithm of x is normally distributed with the same parameters. The probability density function is the following:

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, x > 0$$

Limpert et al. [41], and [40] argue that "[s]kewed distributions are particularly common when mean values are low, variances are large, and values cannot be negative". For the difference between Lognormal and normal variability, they say "[a] major difference ... is

8

that the effects can be additive or multiplicative, thus leading to normal or lognormal distributions". Limpert et al. [41], as discussed in the Introduction, show how the Lognormal distribution is generated from a sequence of multiplicative random effects, $x \times c$ and $\frac{x}{c}$, where x is a given value of the distribution and c any constant, equivalent to the additive random effects that generate a normal distribution, $x + c$, and $x - c$. They add that, as in the normal distribution, an additive 68–95–99.7 rule applies (i.e., the arithmetic value plus and minus one, two, or three standard deviations), so also in the Lognormal distribution a multiplicative rule 68–95–99.7 rule emerges for iterative multiplication and division by standard deviations.

3.1 Testing the tails: Pareto distribution or Lognormal?

3.1.1 The u cut-off parameter

The first step to test the behaviour at the tails is to establish at what point each tail starts. To the left of the u parameter the behaviour is assumed to be Lognormal. To the right there is competing behaviour between the Pareto and the Lognormal. To establish the u parameter we use the Clauset et al. methodology described in [12] and [13]: they choose the u that minimises the distance between the Power Law model and the empirical data. To measure the distance they use the maximum distance between the cumulative distribution function (CDF):

$$D = max_{x \geq u}|S(x) - P(x)|$$

$S(x)$ is the CDF of the data for the observations with value at least u, and $P(x)$ is the CDF of the Power-law model that best fits the data in the region $x \geq u$. u is then the value that minimises D.[13] Clauset et al. use the *max* or *infinity* metric, d_∞, i.e, the so-called Chebyshev distance in the space of functions. They suggest that this Kolmogorov-Smirnov (KS) statistic is well behaved for this model. They even show that this method performs better than a Bayesian information criterion (BIC) another alternative they im-

plement. Malevergne et al. [43] propose a Maximum-likelihood estimation (MLE) method maximising the likelihood of a piecewise-function of a Lognormal below u and a Pareto at and above u. We tested our results against this methodology with similar u estimates. Bee et al. [8] add a ME methods, that is the best approximating density in a non-parametric setup. Generalising it can be said that what is needed is a point estimation method to infer the most statistically plausible estimate that accommodates the Lognormal model to the left and the Pareto one to the right of the estimate.

After calculating u, it is direct to use another MLE estimation to fit the Pareto tail to find the α parameter. Finally, using a non-parametric bootstraping procedure, i.e. by repeatedly subsampling from the empirical distributions, we can find 95% confidence intervals for each described parameter. Figures 6 and 8 show the fitted CCDF distributions with a dashed line going from the estimated u throughout the tail with a slope equal to the estimated α for all the distributions.

We use different methods to implement each statistic including code by Clauset [11] for MATLAB [58], Alstott et al. [2] for Python [59], and Gillespie [26] and Shalizi [55] for R [48], and our own code to implement the methods not readily available. Our results for the number of firms and the total employees at the u parameter for the eight distributions under analysis are presented in Table 5 and the 95% confidence intervals in Table 6 and Figures 9 and 10. It can be seen from the results that both reported methods, Clauset et al. and Gillespie, provide the same outcomes: for the small sample frequency distributions, the $\hat{u}$ cut-off includes the whole distribution (but for 1851 this starts at a $\hat{u}$ of three for the frequency of firms), while for the large sample size distributions the cut-off is evenly at a $\hat{u}$ of ten employees. The $\hat{\alpha}$ estimations of the shape parameter are also given. In Table 6, it is also shown the number of observations within the tail or $Total\ obs - rank(\hat{u})$ and the percentage of firms or employees respectively in the frequency and the size tails, which range from 53% to 100% for the frequency tail and from 60% to 77% for the size tail.[4]

---

[4]We have tested all our methods by replicating the well-known US Census cities data in Clauset et al. dataset [13] to ensure that the calculation methods are equivalent. We also used synthetic Power Law distributions to test the various tests generated by means of the following random number generator for Power Law stated by Clauset et al. [13]: $\left\lfloor (x_{min} - \frac{1}{2})(1 - r)^{-\frac{1}{(\alpha-1)}} + \frac{1}{2} \right\rfloor$ where r is a random real number uniformly distributed in the interval $[0, 1)$ and $\lfloor\ \rfloor$ is the floor function.

3.1.2 A proper Pareto test

Hypothesis testing, following Casella and Berger [9], is an inferential method where a statement about a population parameter is assessed to decide based on a sample of the population, which of two complementary hypotheses is true: the *null*, H, and the *alternative* hypotheses, K. A test is a function of the sample $T(x_1, x_2, ..., x_n) = T(\mathbf{X})$ which specifies a region on the parameter space Θ where the *null* hypothesis is accepted, Θ_H, and a rejection region where the *alternative* is accepted, Θ_K.

Clauset et al. [13] present a proper Pareto test claiming that "[m]ost previous empirical studies of ostensibly power-law [they use power-law but in practice they refer to Pareto] distributed data have not attempted to test the power-law hypothesis quantitatively." They introduce a *null* hypothesis of a Pareto distribution versus an *alternative* hypothesis that a Pareto distribution is not plausible. The method is straightforward. First they fit the empirical data to a Pareto model using MLE and then calculate the same KS statistic they used to pin down the u parameter. They then generate a vast number of Pareto synthetic distributions using as parameters the ones achieved in the first step and calculate KS for these models. Then the proportion of these values larger than the empirical data is the p-value. Figure 11 depicts the p-value for this test of the null hypothesis that the upper tail of our frequency distributions is Pareto using the methodology of Clauset et al. [13]. The test is positive to the null for all the years except 1871.

3.1.3 Pareto or Lognormal?

When making a decision (see, for example, Lehmann and Romano [39]), it is necessary to distinguish Type I error (rejecting H when H is true) from Type II error (accepting H when K is true). The power function $\beta(\theta)$—where θ is the parameter tested—contains the information to test these. The ideal is to have $\beta(\theta) = 0$ for values of the parameter $\theta \in \Theta_H$ and $\beta(\theta) = 1$ for values of the parameter $\theta \in \Theta_H$. A test is assessed according to the degree its accomplishes this ideal. Normally, Type 1 error is fixed at a given level of significance equal to 0.01, 0.05 or 0.1. Moreover, a test is *uniformly most powerful* (UMP) if its power

11

function is closer to one than any other test testing the same H and K hypotheses for every value of the parameter in the K subspace. This means, that after setting a Type 1 error at a given level of significance, we are trying to find the best test among the tests of the same class comparing its performance *not to commit a Type 2 error*. Finally, when this is not achievable, it can be demonstrated that sometimes an *unbiased* UMP, or UMPU, test can be found, where unbiased simply means that its probability of committing Type 2 error is always higher in the relevant hypothesis subspace than the probability of committing Type 1 error, which is a generally plausible assumption.

After introducing this approach, we can present the H and K hypotheses we are confronting as summarised by Malevergne et al. [43]:

"H: Pareto distribution for values of x larger than some threshold u and

K: Lognormal distribution also for values of x above the same threshold u"[43]

To test these hypotheses, Malevergne et al. show that an UMP test is not achievable, while they suggest as an UMPU test the maximum likelihood ratio, that is "with insertion in the maximum likelihood ratio of the maximum likelihood estimates of the unknown parameters instead of the true values".[43] In mathematical statistics, this test is called the Wilks test [43]. Malevergne et al. show that an equivalent statistic to this UMPU-Wilks test is the clipped sample coefficient of variation $\hat{c} = min(1, c)$, where c is the sample coefficient of variation, that is the ratio of the sample standard deviation to the sample mean:

$$c = \frac{s_X}{\overline{X}}$$

This test proposed by Malevergne et al. overcomes the biggest deficiency of competing tests previously used in the literature that have very limited power such as the *L-test* or χ^2*-test* used in the famous Eeckhout Figure 2 (see [16]) where the confidence interval under the null of Lognormal behaviour expands disproportionately at the tail of the distribution completely compromising the test ability to discern between the vying hypotheses. According to Malevergne et al. the clipped sample coefficient of variation can be deemed as a sufficient statistic for the optimal UMPU test. The critical function of this test is a function

of the sample $T(x_1, x_2, ..., x_n) = T(\mathbf{X})$ equal to, as shown by [43]:

$$T(\mathbf{X}) = 1 \quad if \quad \hat{c} < h$$

$$T(\mathbf{X}) = 0 \quad if \quad \hat{c} \geq h$$

where h is the critical value chosen "such that the probability that the inequality $\hat{c} < h$ holds under the null hypothesis [Pareto distribution] is equal to the (small) preliminary chosen [level of significance]".[43] To derive h, Malevergne et al. suggest two methods: a saddle point method as shown by Del Castillo and Puig [15] and Gatto and Jammalamadaka [24] and a simpler Monte Carlo approximation. Malevergne et al. refer to Del Castillo and Puig [15] who showed that for the problem of testing the null hypothesis that the upper tail is exponential against the alternative of a truncated normal, the UMPU-Wilks test is, also, the clipped sample coefficient of variation. Finally, they describe how to perform the simpler Monte Carlo method to find h. For this a large enough number of samples M of the standard exponential function with unit parameter has to be generated—they suggest at least 10,000—. Then, for each sample, $\hat{c}$ is calculated and compared with that of the sample under scrutiny. Therefore, as argued in [43], the "fraction of exceedances provides a good statistical estimate of the corresponding p-value of the null hypothesis [Pareto distribution]".

Yet another statistic is the logarithm R of the ratio of the likelihoods of the data under two competing distributions presented by Clauset et al. [13]. If the outcome of the test is positive then the first model has preeminence over the second, if the outcome is negative the second should prevail. But it is also necessary to correct for random fluctuations. To do so the ratio R is standardised by the standard deviation σ using the method first presented by Vuong [60]. The method gives a p-value that tells "whether the observed sign of R is statistically significant".[13] If the p-value is small, say below a significance level of 0.01 the test is unlikely to be a chance result of fluctuations. The hypotheses tested by Clauset et al. [13] are the following:

H: Pareto distribution for values of x larger than some threshold u and

K_1: exponential distribution also for values of x above the same threshold

K_2: stretched exponential distribution also for values of x above the same threshold

K_3: Lognormal distribution also for values of x above the same threshold

The results comparing the Pareto distribution with the exponential—which is the absolute minimum alternative because the usual definition of "heavy-tail" is against the "light-tailed" exponential behaviour [4]—the stretched exponential, and the Lognormal are given in Table 7. All the distributions are confirmed as "heavy-tailed" because the likelihood ratios are positive and the p-values smaller than 0.01. For the stretched exponential the frequency distributions for years 1851 and 1871 are positive and significant, but the other two frequency distributions for 1861 and 1881 are not, even though they are marginally non-significant. The four size distributions compared to the stretched exponential have best fit by the Pareto distribution with strongly significant p-values. The comparison between Pareto and Lognormal for the size distributions, have the year 1851 significant at a 0.05 level and the year 1861 at any level, ruling out the Lognormal model and confirming the Pareto behaviour. For 1871, the significance is slightly inconclusive and for the year 1881, the significance is clearly inconclusive. With these results it is possible to answer the main question of this paper: at least for the years 1851 and 1861 the size distributions have better fit by a Pareto tail than by a Lognormal one, defining the tails as all the observations above the estimated u cut-off parameter. The test is inconclusive for the frequency distributions with high p-values and alternating signs. Thus, it is not possible to rule out the Lognormal or the Pareto distributions with the methods employed for the frequency distributions.

3.2 To Zipf or not to Zipf?

Having constructed, in Table 6, the 95% confidence intervals for the α parameters for the two tails, size and frequency, it is direct to test the null that the model follows a Zipf behaviour. This is relevant, as noted earlier, where Zipf behaviour is a "Pareto behaviour with an exponent equal to 1" or a "Power Law with a cut-off and an exponent equal to 1". The answer is an absolute no. None of the intervals includes a value compatible with Zipf behaviour in the tail. So a rigorous and simple hypothesis testing permits us to be

conclusive in this respect that the two tails are not Zipfian at the 0.05 level as implied by the confidence intervals displayed in Table 6.

4   A comparison with modern times

We compare our data for the latest year (1881) with figures taken from Robson and Gallagher [51] for the UK for 1987, as shown in Table 8. This is the closest year of modern data to the historical censuses 1851-81. In the table, we compare the number of firms, the number of employees and the percentage of employment (the percentage of employees in a given bin out of the total employees in the economy) for size category used by Robson and Gallagher for 1987 with our data for 1881. In this table we have to use bins to match those in the Robson and Gallagher data, although we acknowledge that bins distort the data, as noted earlier. Unfortunately, we can not perform here the same tests developed in the previous sections of the paper because the full dataset of Robson and Gallagher [51] is not archived. If data reemerged it would be a promising area for future research. Also, it needs to be stressed that UK data are not equivalent to our data for England and Wales. Nevertheless, the comparisons are valuable. There is an increase of 399.7% (5.8% annually) in the number of firms, from 181,125 to 905,155; and in the number of employees of 574.1% (6.2% annually), from 2,142,404 to 14,433,000. In the twenty bin size categories, fifteen show a decrease in the percentage change in percentage employment with a minimum of -63.9% in the 500-999 bin, while five show an increase with a maximum of infinity in the over 10,000 bin because there are no firms in this bin for 1881. There are only 179 very large firms over 10,000 in 1987 and they account for a huge 22% of the employment; in our historical data the largest 179 firms had 12.3% of employment (calculated from our individual firm data), roughly half the modern figure. Although there is variation between bin categories, the dominant pattern is for positive and high percentage increases in the two largest bins, and decreases in all but one of the small and medium-sized categories. Although constrained by differences in the bases of the data over time and the bin categories, the lengthening and fattening of the tail is a continuation of the trend we observe in 1851-81, and in line with the expectations of the Pareto and Lognormal distributions: i.e. the tail increases massively in fatness and length over time. Also, this is in line with economic literature that has suggested that significant concentration into larger firms probably began in the mid-Victorian period after the 1870s and 1880s and continued thereafter (e.g. Hannah [30]).

5   Discussion and conclusions

This paper explores two distributions of firm data (size and frequency) during the mid-Victorian era to test the Power Law compared to the Lognormal hypotheses at the tails: that is, a tale of two tails. We show that the Pareto distribution fits the tails of the size data better than a Lognormal one for at least the years 1851 and 1861, and also with marginally non-significant values in 1871, but for the frequency distributions the test is inconclusive. Thus, the main contribution of this paper to the literature is to demonstrate that the new dataset for firms size in mid-Victorian era is heavy-tailed, with the tail belonging to the Power Laws rather than the Lognormal distribution. Comparison with more modern data for 1987, although constrained by the available data and categories, confirms that tails have grown larger over time, whilst the main part of the distribution for small firms has been reduced. Also, we have been able to rule out a Zipf behaviour in the tails of our size and frequency distributions. Some of the techniques revisited are the UMPU-Wilks clipped sample coefficient of variation for distinguishing between the Power Laws and Lognormal behaviour, MLE to point estimate the u and the α parameters, KS to measure the distance between the actual data and the fitted one, and log likelihood ratios to compare distributions and Vuong procedure to obtain the p-values for this comparison.

We use the original firm size data at individual level which is exhaustive, thus meeting Fujiwara et al. [18] condition that the data are complete to allow identification of tail effects. And we avoid grouping (binning), which Bee et al. [8] argue is one of the most overlooked limitations of many previous analyses of firm size. Despite, Fujita et al. [17] questioning how to distinguish between distributions: "Is this a solution to the riddle? Our view is that it is ingenious but not entirely satisfactory", we are now in a position to answer for the distributions in the tail. Using the advances in the literature by Clauset et al. [13] and Malevergne et al. [43] we have shown how rigorous assessments of firm size tail behaviour can be used to test between these two tales. With these procedures it is possible to answer the main question of this paper: at least for the years 1851 and 1861 the size distributions have better fit by a Pareto tail than by a Lognormal one, defining the tails as all the observations above the estimated u cut-off parameter. This outcome is conclusive and stands as a new addition to the profuse literature on this issue. We are convinced that these new procedures recently discussed in the literature permits to properly test the tale of the tails:

16

are they Pareto or Lognormal above the cut-off? We expect this discussion to carry on in the literature as more tails, current and past, being questioned on their tail behaviour. At least, we deliver the answer for this newly available mid-Victorian firm size and frequency.

References

[1] Paul Allanson. Farm Size Structure in England and Wales 1939-89. *Journal of Agricultural Economics*, 43(2):137–148, may 1992.

[2] Jeff Alstott, Ed Bullmore, and Dietmar Plenz. Powerlaw: A python package for analysis of heavy-tailed distributions. *PLoS ONE*, 9(1):e85777, 2014.

[3] Sidra Arshad, Shougeng Hu, and Badar Nadeem. Zipf's law and city size distribution: A survey of the literature and future research agenda. *Physica A: Statistical Mechanics and its Applications*, 492:75–92, 2018.

[4] Søren Asmussen. *Applied probability and queues. 2nd ed.* Applications of mathematics; 51. Springer, Berlin; Heidelberg, 2003.

[5] Robert L. Axtell. Zipf Distribution of US Firm Sizes. *Science, New Series*, 293(5536):1818–1820, 2011.

[6] Michael Batty. Rank clocks. *Nature*, 444(7119):592–596, 2006.

[7] Marco Bee, Massimo Riccaboni, and Stefano Schiavo. The size distribution of US cities: Not Pareto, even in the tail. *Economics Letters*, 120(2):232–237, 2013.

[8] Marco Bee, Massimo Riccaboni, and Stefano Schiavo. Where Gibrat meets Zipf: Scale and scope of French firms. *Physica A: Statistical Mechanics and its Applications*, 481:265–275, 2017.

[9] George Casella and Roger L. Berger. *Statistical inference. 2nd ed.*, volume 1. Duxbury, Pacific Gove, CA, 2002.

[10] Pasquale Cirillo. Some evidence about the evolution of the size distribution of Italian firms by age. *Economics Bulletin*, 29(3):1723–1730, 2009.

[11] Aaron Clauset. Powerlaw: Matlab codes for analysis of heavy-tailed distributions, 2009.

[12] Aaron Clauset, Kristian Skrede Gleditsch, and Maxwell Young. On the Frequency of Severe Terrorist Events. *Journal of Conflict Resolution*, 51(1):58–87, 2007.

[13] Aaron Clauset, Cosma Rohilla Shalizi, and M. E. J. Newman. Power-Law Distributions in Empirical Data. *SIAM Review*, 51(4):661–703, 2009.

[14] Bernat Corominas-Murtra and Ricard Sole. Universality of Zipf's Law. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 82(011102):1–9, 2010.

[15] Joan del Castillo and Pedro Puig. The Best Test of Exponentiality against Singly Truncated Normal Alternatives. *Journal of the American Statistical Association*, 94(446):529–532, 1999.

[16] Jan Eeckhout. Gibrat' s Law for (All) Cities. *The American Economic Review*, 99(4):1676–1683, 2009.

[17] Masahisa Fujita, Paul R. Krugman, and Anthony Venables. *The spatial economy cities, regions and international trade.* MIT Press, Cambridge, MA, 2000.

[18] Yoshi Fujiwara, Corrado Di Guilmi, Hideaki Aoyama, Mauro Gallegati, and Wataru Souma. Do Pareto-Zipf and Gibrat laws hold true? An analysis with European firms. *Physica A: Statistical Mechanics and its Applications*, 335(1-2):197–216, 2004.

[19] Xavier Gabaix. Zipf' s Law and the Growth of Cities. *The American Economic Review*, 89(2), 1999.

[20] Xavier Gabaix. Zipf's Law for cities: An explanation. *The Quarterly Journal of Economics*, (August):739–767, 1999.

[21] Edoardo Gaffeo, Corrado Di Guilmi, Mauro Gallegati, and Alberto Russo. On the mean/variance relationship of the firm size distribution: Evidence and some theory. *Ecological Complexity*, 11:109–117, 2012.

[22] Edoardo Gaffeo, Mauro Gallegati, and Antonio Palestrini. On the size distribution of firms: Additional evidence from the G7 countries. *Physica A: Statistical Mechanics and its Applications*, 324(1-2):117–123, 2003.

[23] Li Gan, Dong Li, and Shunfeng Song. Is the Zipf law spurious in explaining city-size distributions? *Economics Letters*, 92(2):256–262, 2006.

[24] Riccardo Gatto and S. Rao Jammalamadaka. A saddlepoint approximation for testing exponentiality against some increasing failure rate alternatives. *Statistics & Probability Letters*, 58:71–81, 2002.

[25] Robert Gibrat. *Les inegalites economiques*. Sirey, Paris, 1931.

[26] Colin S. Gillespie. Fitting heavy tailed distributions: the poweRlaw package. *Journal of Statistical Software*, 64(2), 2015.

[27] David Grigg. Farm size in England and Wales, from early Victorian times to the present. *Agricultural History Review*, 35(2):179–189, 1987.

[28] Jan A. Groenewald. Returns to size and structure of agriculture: A suggested interpretation. *Development Southern Africa*, 8(3):329–342, 1991.

[29] Alice Payne Hackett. *70 Years of best sellers, 1895-1965*. R. R. Bowker Co., New York, 1967.

[30] Leslie Hannah. *The rise of the corporate economy. 2nd ed.* Methuen, London, 1983.

[31] Torsten Heinrich and Shuanping Dai. Diversity of firm sizes, complexity, and industry structure in the Chinese economy. *Structural Change and Economic Dynamics*, 37:90–106, 2016.

[32] Edward Higgs and Kevin Schürer. Integrated Census Microdata (I-CeM), 1851-1911, UK Data Archive data deposit SN-7481, 2014.

[33] Sang Hoon, Zhuhua Jiang, Chongcheul Cheong, and Seong-min Yoon. Changes of firm size distribution: The case of Korea. *Physica A: Statistical Mechanics and its Applications*, 390(2):319–327, 2011.

[34] Mikael Huss and Petter Holme. Currency and commodity metabolites: Their identification and relation to the modularity of metabolic networks. *IET Systems Biology*, 1(5):280–5, 2006.

[35] Takashi Ito, Kosuke Tashiro, Shigeru Muta, Ritsuko Ozawa, Tomoko Chiba, Mayumi Nishizawa, Kiyoshi Yamamoto, Satoru Kuhara, and Yoshiyuki Sakaki. Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proceedings of the National Academy of Sciences*, 97(3):1143–1147, 2000.

[36] N.L. Johnson, S Kotz, and N. Balakrishnan. *Continuous Univariate Distributions,*, volume 1. Wiley, New York, 1970.

[37] Jovan Karamata. Sur un mode de croissance régulière des fonctions. *Mathematica (Cluj)*, 4:38–53, 1930.

[38] Jovan Karamata. Sur un mode de croissance régulière. Théorèmes fondamentaux. *Bulletin de la Société Mathématique de France*, 61:55–62, 1933.

[39] Erich Leo Lehmann and Joseph Romano. *Testing statistical hypotheses.* Springer texts in statistics. Springer, New York, N.Y., 3rd ed. edition, 2008.

[40] Eckhard Limpert and Werner A. Stahel. What is the normal distribution? *Significance*, (February):8–9, 2017.

[41] Eckhard Limpert, Werner A. Stahel, and Markus Abbt. Log-normal Distributions. *BioScience*, 51(5):341–352, 2001.

[42] Philip Lund and Roger Price. The measurement of average farm size. *Journal of Agricultural Economics*, 49(1):100–110, 1998.

[43] Yannick Malevergne, Vladilen Pisarenko, and Didier Sornette. Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 83(3):1–11, 2011.

[44] Michael Mitzenmacher. A Brief History of Lognormal and Power Law Distributions. *Internet Mathematics*, 2018.

[45] M. E.J. Newman. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5):323–351, 2005.

[46] Kenji Okuyama, Misako Takayasu, and Hideki Takayasu. Zipf's law in income distribution of companies. *Physica A: Statistical Mechanics and its Applications*, 269(1):125–131, 1999.

[47] Rui Pascoal, Mário Augusto, and A. M. Monteiro. Size distribution of Portuguese firms between 2006 and 2012. *Physica A: Statistical Mechanics and its Applications*, 458:342–355, 2016.

[48] R Core Team. R: A Language and Environment for Statistical Computing. Computer software, 2017.

[49] Sidney Redner. How popular is your paper? An empirical study of the citation distribution. *European Physical Journal B*, 4(2):131–134, 1998.

[50] David C. Roberts and Donald L. Turcotte. Fractality and Self-Organized Criticality of Wars. *Fractals*, 06(04):351–357, 1998.

[51] Geoffrey B. Robson and Colin C. Gallagher. Change in the size distribution of UK firms. *Small Business Economics*, 6(4):299–312, 1994.

[52] Alexander I. Saichev, Yannick Malevergne, and Didier Sornette. Theory of Zipf's Law and of General Power Law Distributions with Gibrat's law of Proportional Growth. pages 1–10, 2008.

[53] Alexander I. Saichev, Yannick Malevergne, and Didier Sornette. *Theory of Zipf's law and beyond.* Springer, Heidelberg; London, 2010.

[54] Eugene Seneta. *Regularly Varying Functions*, volume 508 of *Lecture Notes in Mathematics.* Springer, Berlin; Heidelberg, 1976.

[55] Cosma Rohilla Shalizi. R Code for The Estimation of Power Laws and Their Comparison to Heavy-Tailed Alternatives, 2007.

[56] Felisa A. Smith, Rosemary E. Elliott Smith, S. Kathleen Lyons, and Jonathan L Payne. Body size downgrading of mammals over the late Quaternary. *Science*, 360(6386):310–313, 2018.

[57] John Sutton. The variance of firm growth rates: The 'scaling' puzzle. *Physica A: Statistical Mechanics and its Applications*, 312(3-4):577–590, 2002.

[58] The MathWorks Inc. MATLAB. Computer software, 2016.

[59] Guido van Rossum. Python. Computer software, 2017.

[60] Quang H. Vuong. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, 57(2):307–333, 1989.

[61] Walter Willinger and Vern Paxson. Where Mathematics meets the Internet Teletra. *Theory and Internet Engineering*, 45(8):1–12, 1998.

[62] Jianhua Zhang, Qinghua Chen, and Yougui Wang. Zipf distribution in top Chinese firms and an economic explanation. *Physica A: Statistical Mechanics and its Applications*, 388(10):2020–2024, 2009.

[63] George K. Zipf. *Human behaviour and the principle of least effort.* [s.n.], [s.l.], 1949.

Figure 1: Frequency versus size 1851-1881. Double decimal logarithm scale (own data replication of Redner's (1998) Figure 1). (Upper left quadrant 1851, upper right quadrant 1861, lower left quadrant 1871, lower right quadrant 1881).
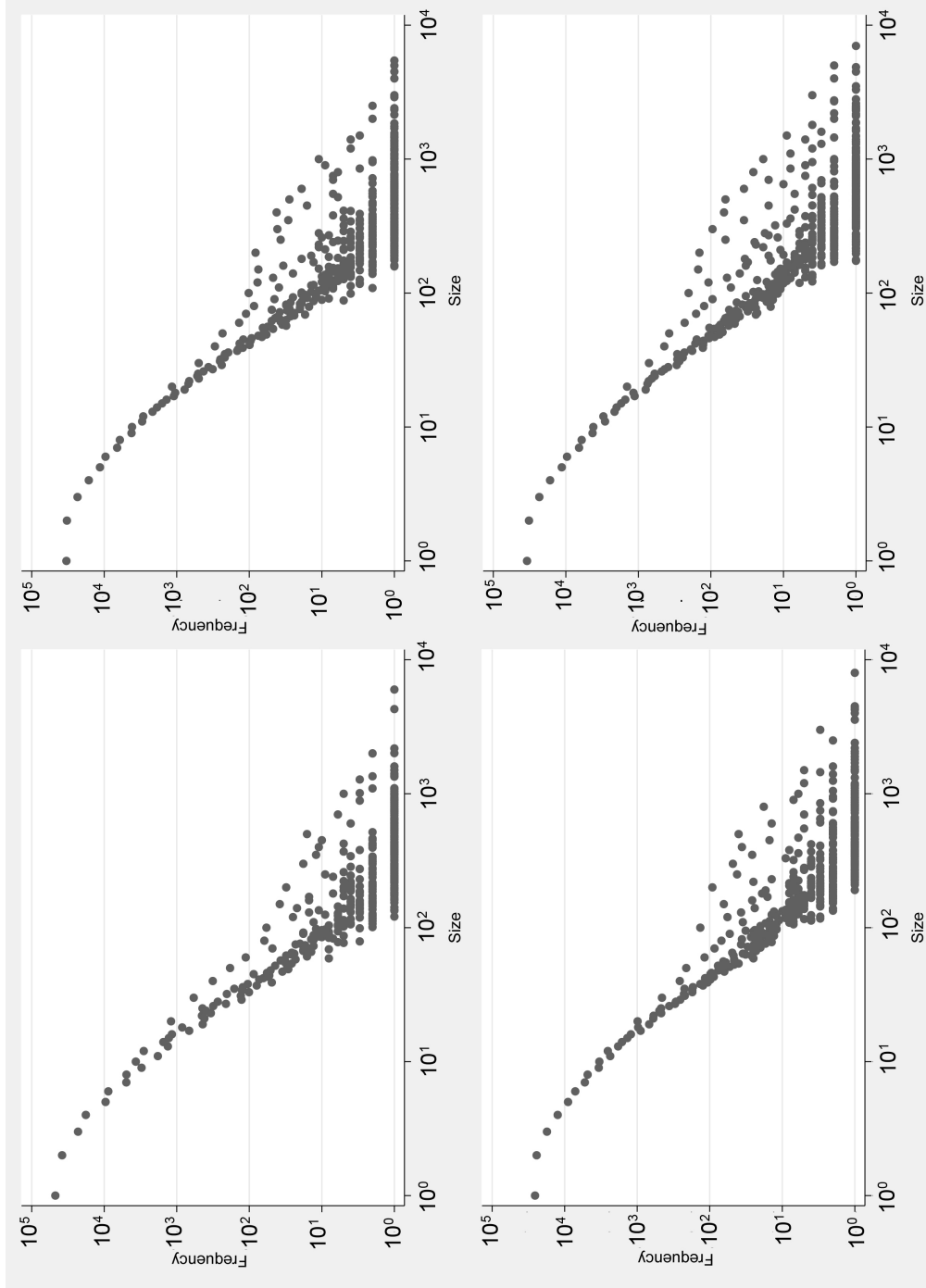
Figure 2: Frequency distribution for Frequency < 40 for the censuses 1851, 1861, 1871, 1881.
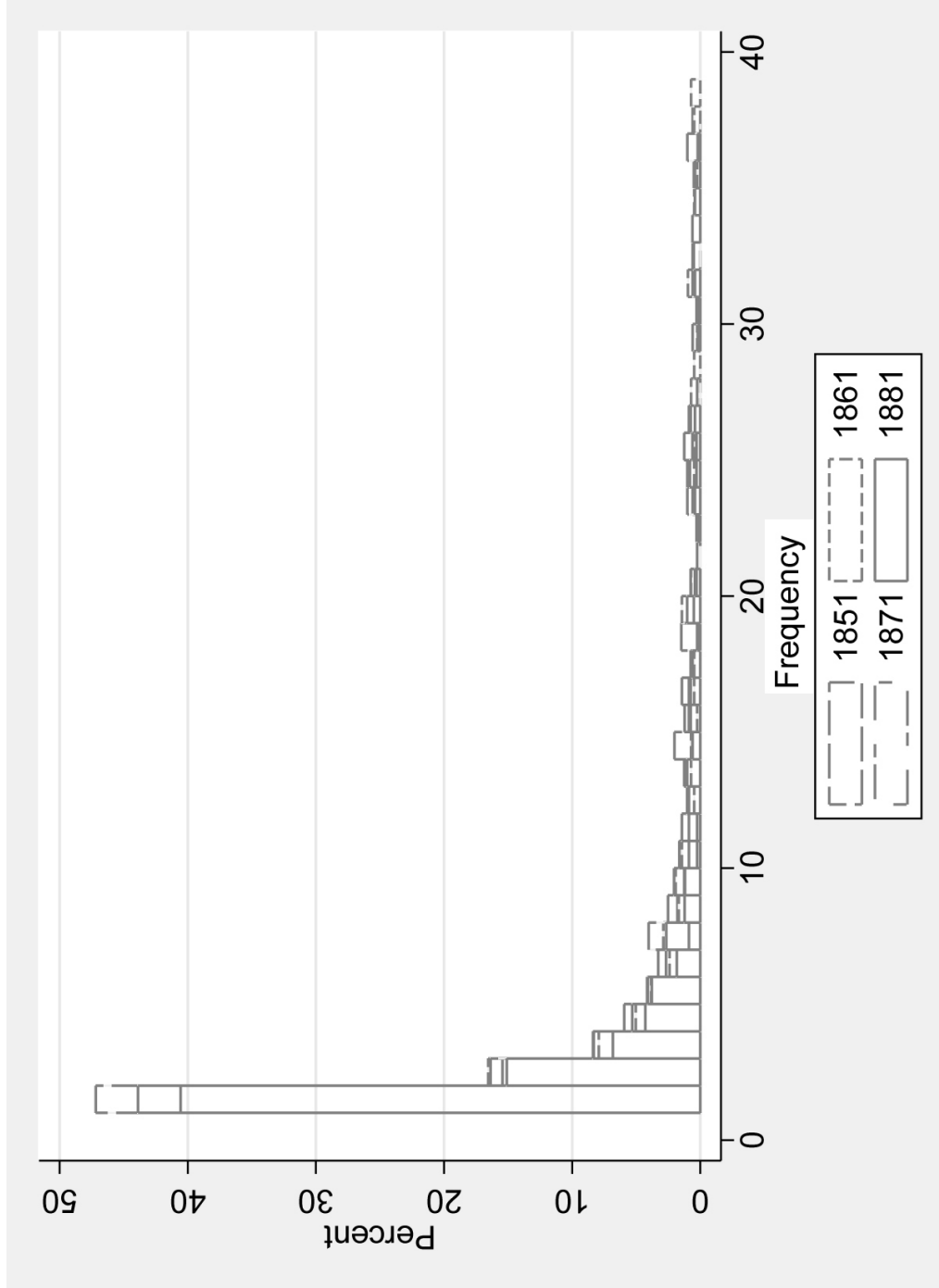
Figure 3: Size distribution for Size < 40 for the censuses 1851, 1861, 1871, 1881.
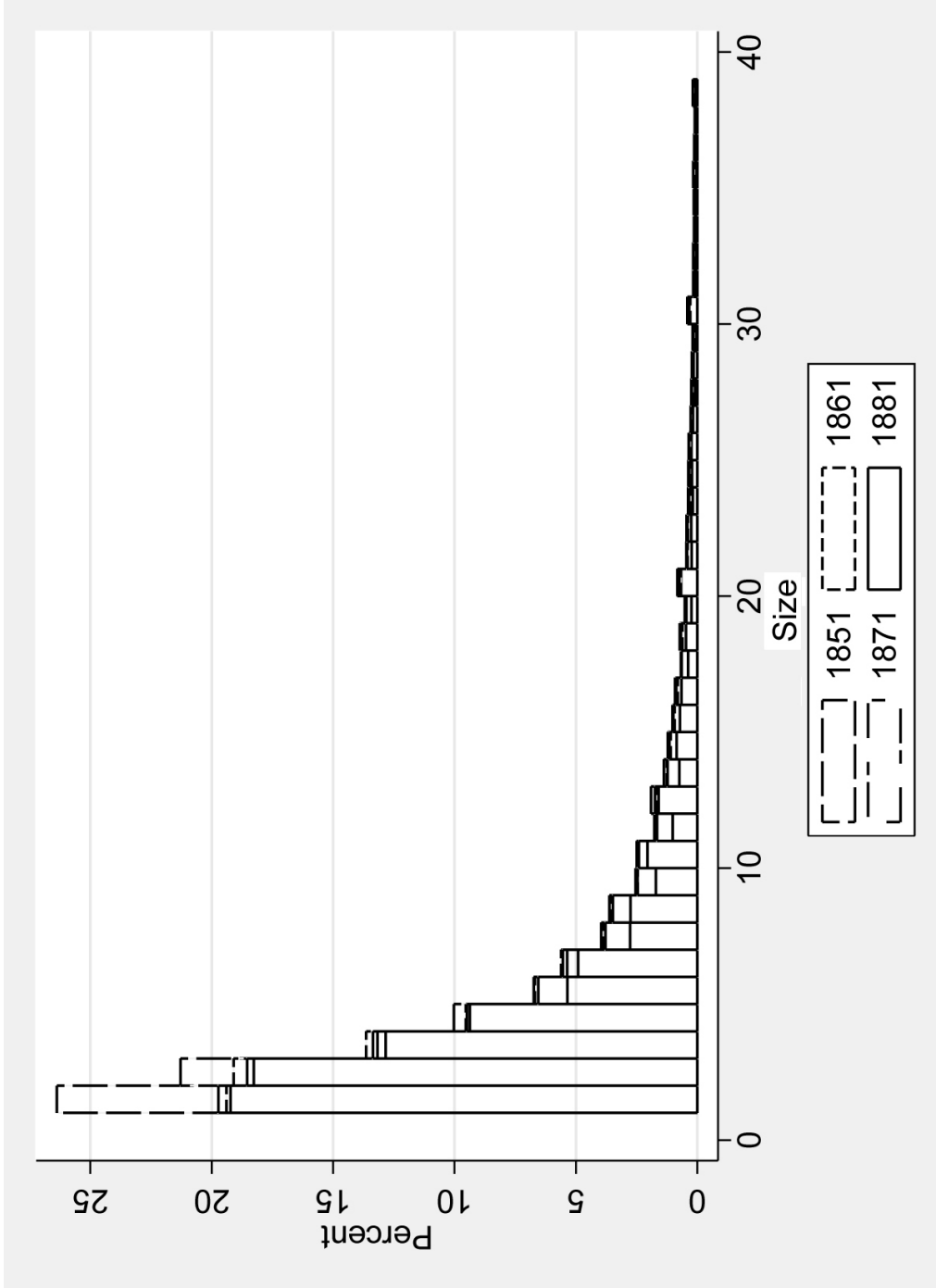
Figure 4: Diagram of the frequency of firms in 1851-1881. (Upper left quadrant 1851, upper right quadrant 1861, lower left quadrant 1871, lower right quadrant 1881).
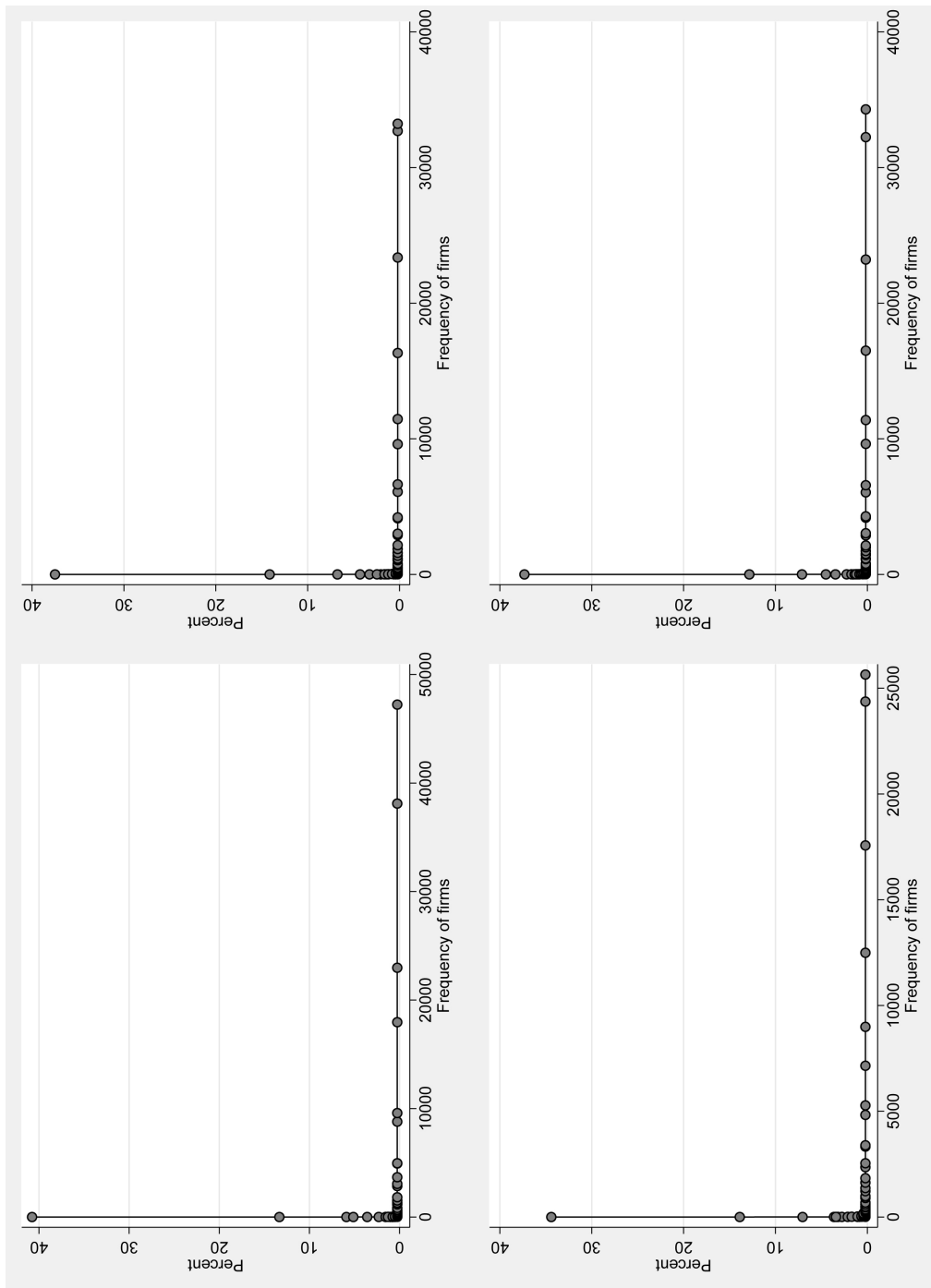
Figure 5: Rank/Frequency plots in decimal logarithmic scales of the frequency of firms by rank of firm size frequency in 1851-1881. (Upper left quadrant 1851, upper right quadrant 1861, lower left quadrant 1871, lower right quadrant 1881).
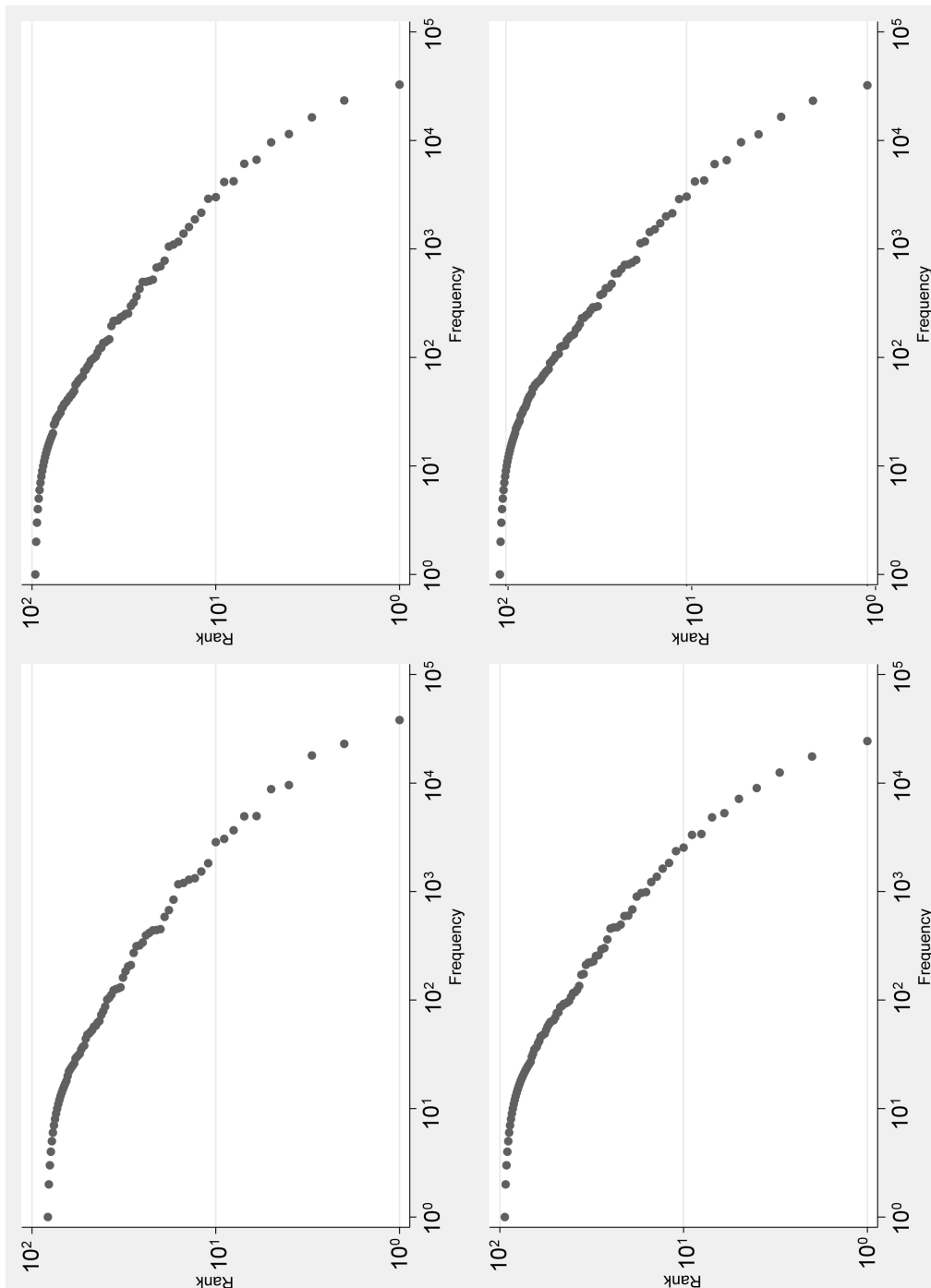
Figure 6: Pareto plots decimal logarithmic scales of the frequency of firms in 1851-1881 using the functions plfit and plplot in the paper of Clauset et al. [13]. x=frequency. (Upper left quadrant 1851, upper right quadrant 1861, lower left quadrant 1871, lower right quadrant 1881).
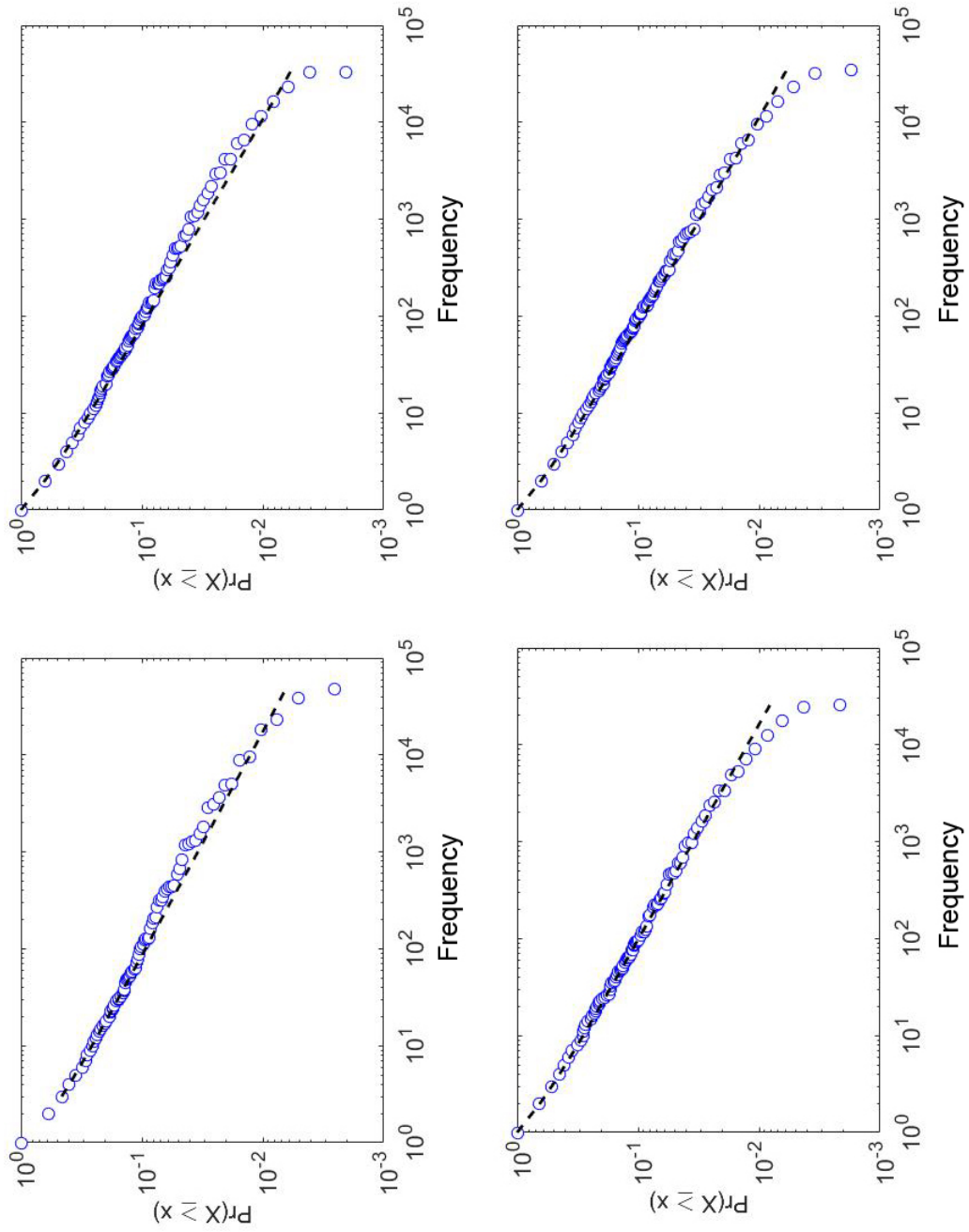
Figure 7: Rank/Size plots in decimal logarithmic scales of rank and size of firms in 1851-1881. (Upper left quadrant 1851, upper right quadrant 1861, lower left quadrant 1871, lower right quadrant 1881).
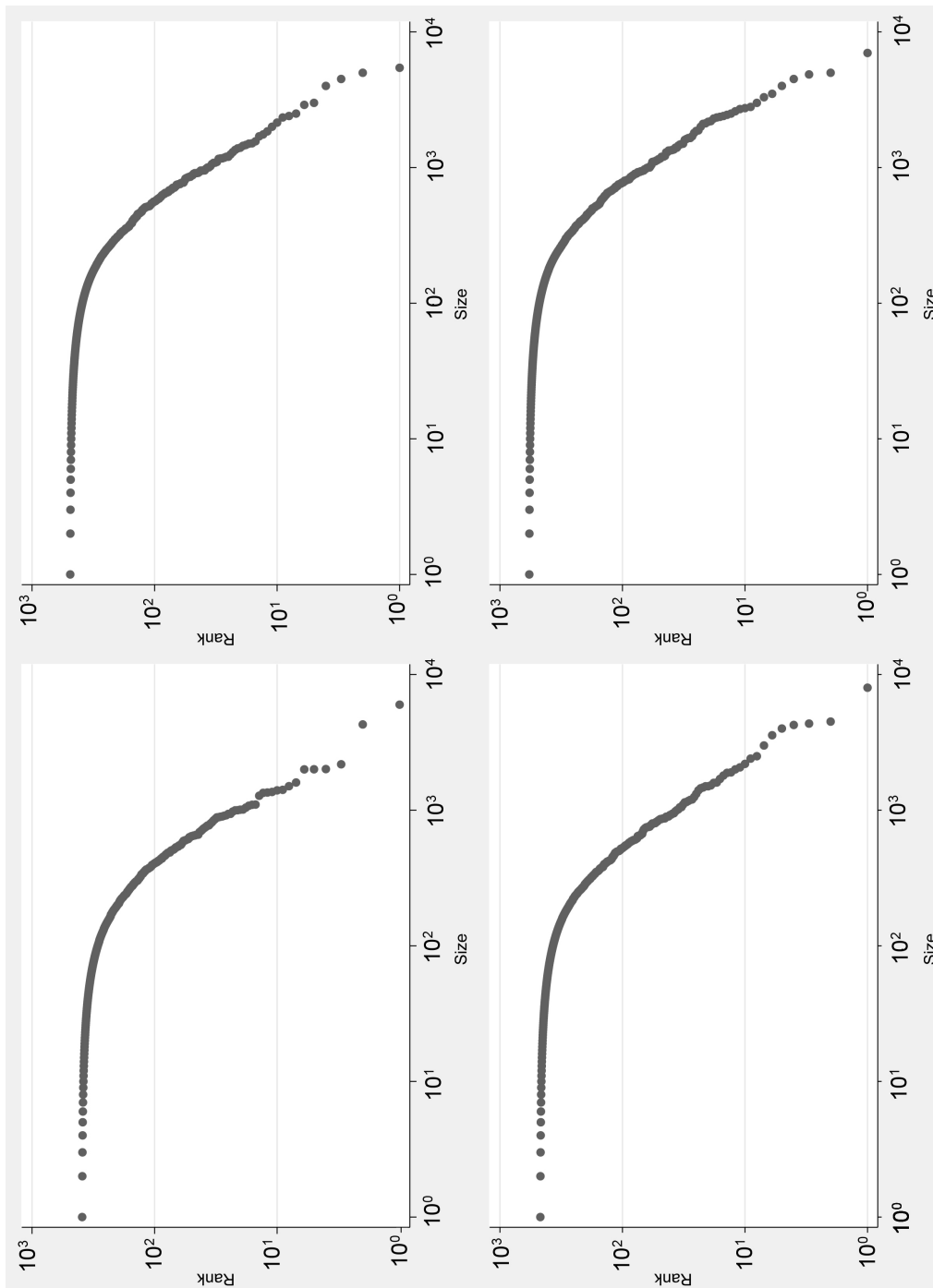
Figure 8: Pareto plots in decimal logarithmic scales of size of firms in 1851-1881 using the functions plfit and plplot in the paper of Clauset et al. [13]. x=size. (Upper left quadrant 1851, upper right quadrant 1861, lower left quadrant 1871, lower right quadrant 1881).
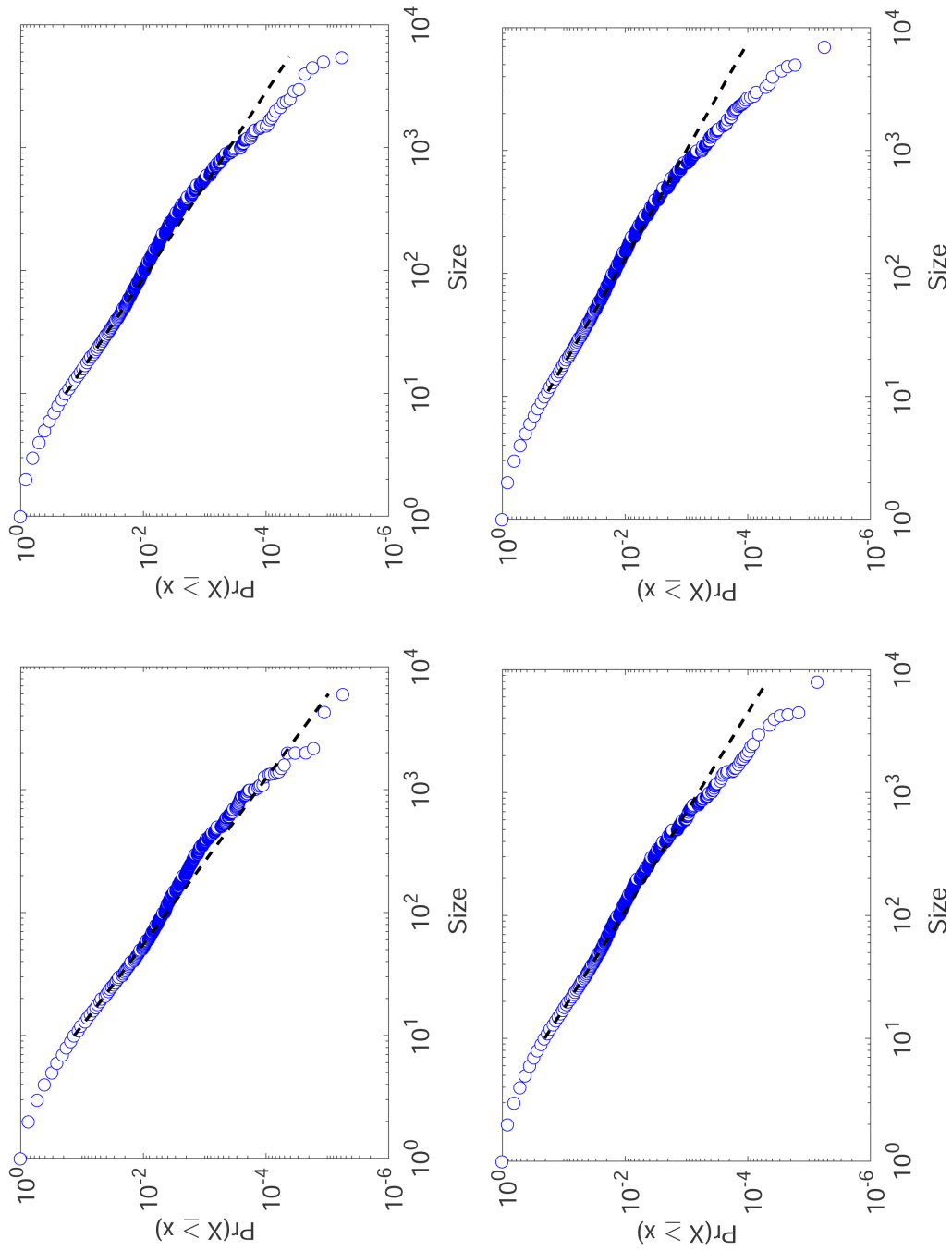
Figure 9: Confidence intervals for the frequency distributions parameters (upper row) and standard deviations of the parameters (lower row) u (left), α (centre), and the tail (right) i.e. $Total\ obs-rank(\hat{u})$. See text for details. (Upper left quadrant 1851, upper right quadrant 1861, lower left quadrant 1871, lower right quadrant 1881). Own implementation of Gillespie [26] method.

31
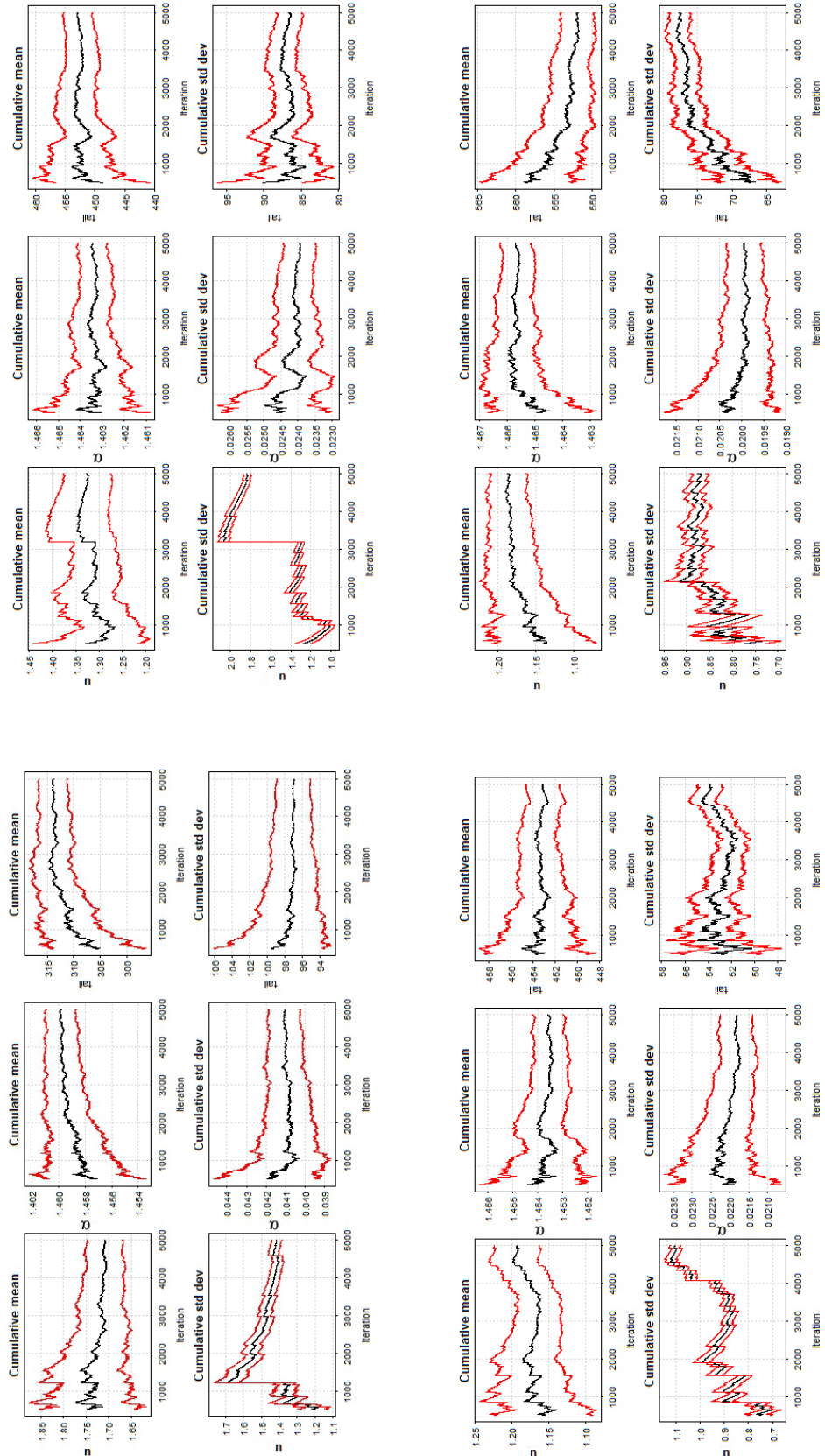
Figure 10: Confidence intervals for the size distributions parameters (upper row) and standard deviations of the parameters (lower row) u (left), α (centre), and the tail (right), i.e. $Total\ obs - rank(\hat{u})$. See text for details. (Upper left quadrant 1851, upper right quadrant 1861, lower left quadrant 1871, lower right quadrant 1881). Own implementation of Gillespie [26] method.
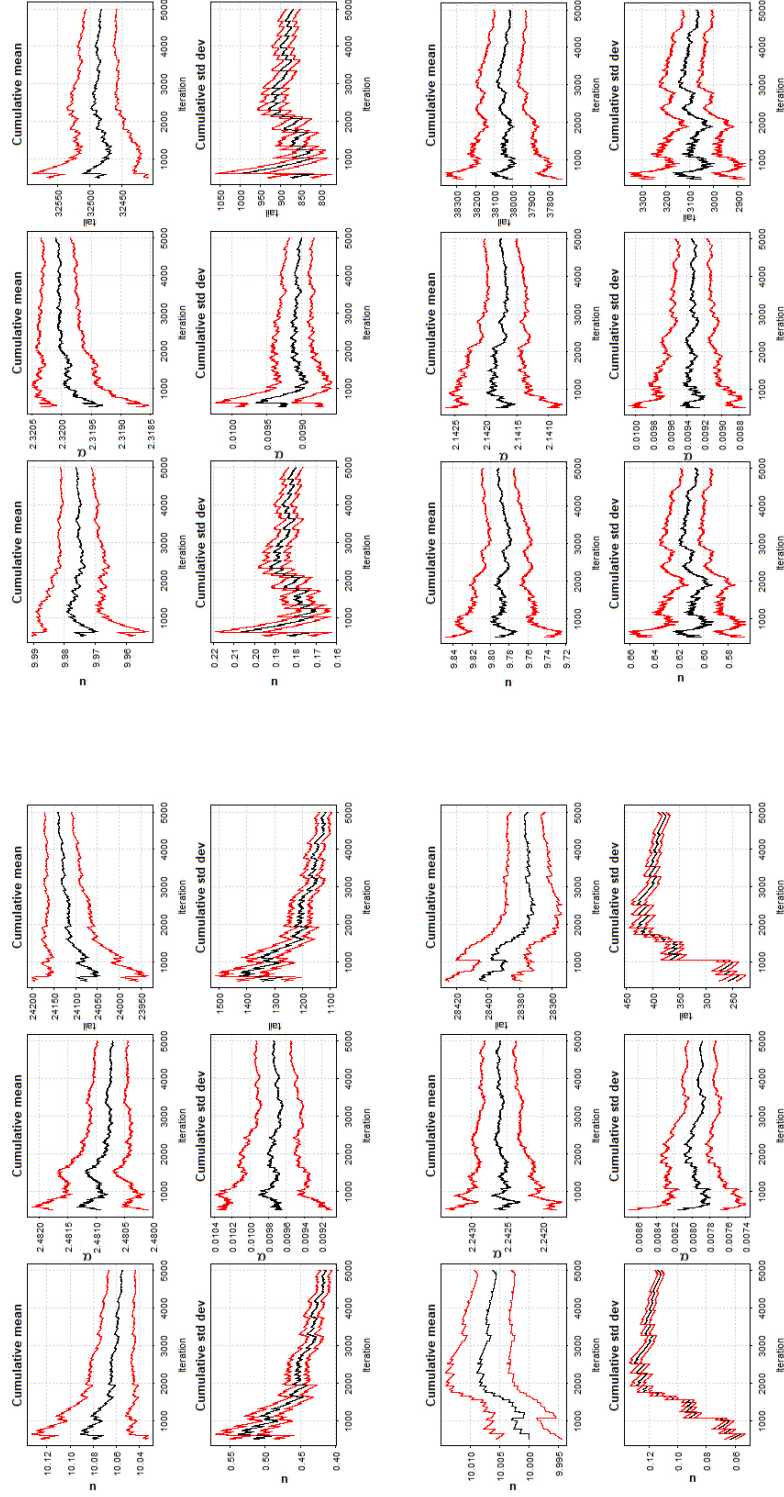
Figure 11: P-values for frequency distributions in Goodness-of-Fit test as described in Clauset et al. [13] Pareto test. See text for details. (Upper left quadrant 1851, upper right quadrant 1861, lower left quadrant 1871, lower right quadrant 1881).

33
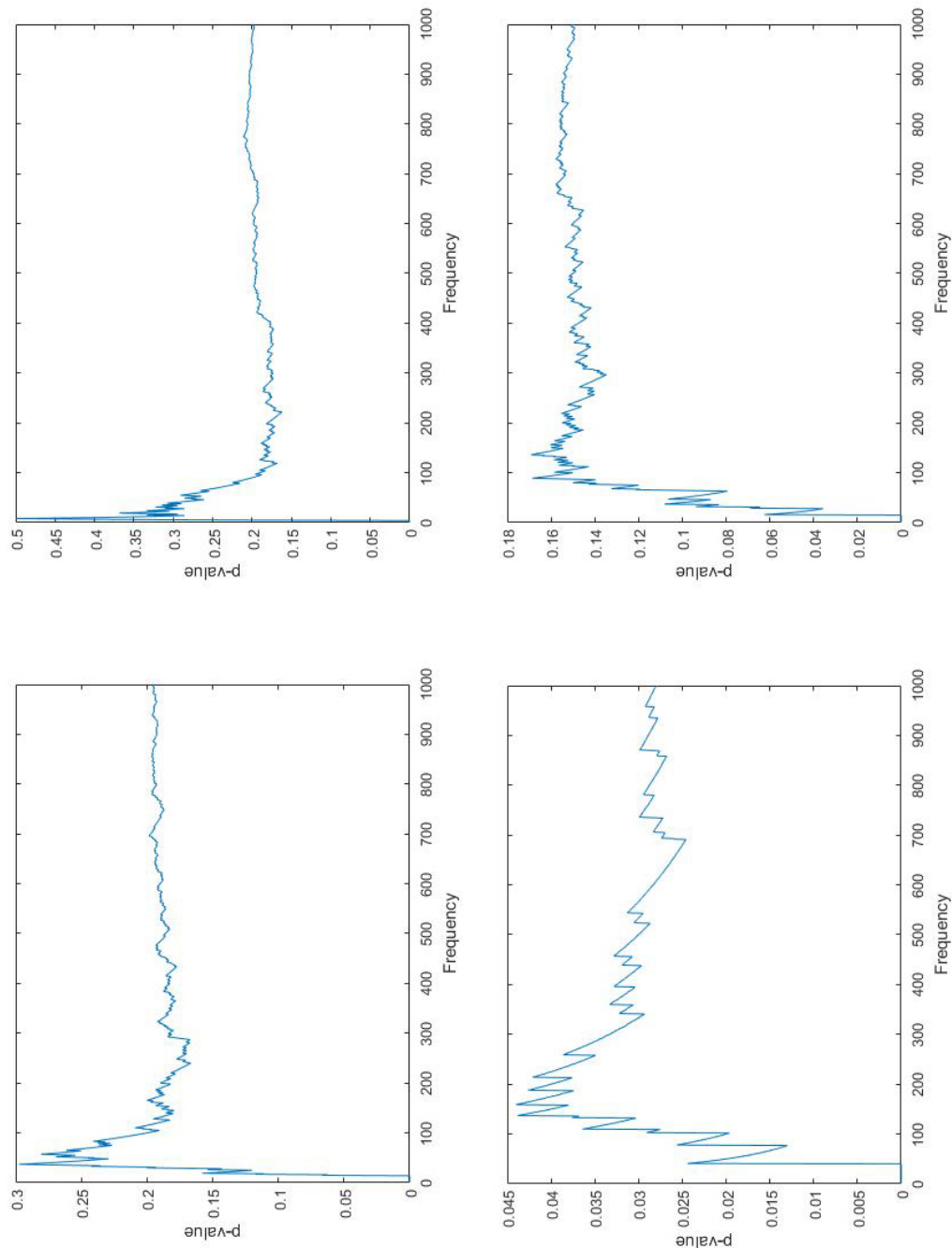
Table 1: The ten most frequent firm sizes in England and Wales for each census year.

Note: F_1/F_i denotes the ratio of each frequency relative to the frequency of the most frequent firm size.

Sources: Censuses 1851-81.

Grey cells, in 1851, show that the function converting rank to size (and its inverse) is not necessarily monotonic.

Rank	Size	Frequency	F_1/F_i	Rank	Size	Frequency	F_1/F_i	Rank	Size	Frequency	F_1/F_i	Rank	Size	Frequency	F_1/F_i
1	1	47,231	1.00	1	1	33,230	1.00	1	1	25,644	1.00	1	1	34,291	1.00
2	2	38,111	1.24	2	2	32,700	1.02	2	2	24,367	1.05	2	2	32,243	1.06
3	3	22,977	2.06	3	3	23,366	1.42	3	3	17,567	1.46	3	3	23,218	1.48
4	4	17,960	2.63	4	4	16,327	2.04	4	4	12,493	2.05	4	4	16,500	2.08
5	5	9,585	4.93	5	5	11,456	2.90	5	5	8,990	2.85	5	5	11,382	3.01
6	6	8,791	5.37	6	6	9,607	3.46	6	6	7,152	3.59	6	6	9,627	3.56
7	7	4,960	9.52	7	7	6,641	5.00	7	7	5,278	4.86	7	7	6,578	5.21
8	8	4,936	9.57	8	8	6,094	5.45	8	8	4,829	5.31	8	8	6,052	5.67
9	10	3,671	12.87	9	9	4,205	7.90	9	9	3,392	7.56	9	9	4,283	8.01
10	9	3,059	15.44	10	10	4,140	8.03	10	10	3,330	7.70	10	10	4,182	8.20
		1851 Census				1861 Census				1871 Census				1881 Census	

Table 2: The ten largest firm sizes in England and Wales for each census year.
Note: F_{last}/F_i denotes the ratio of each frequency relative to the frequency of the least frequent firm size. Sources: Censuses 1851-81.

Rank	Size	Frequency	F_{last}/F_i	Rank	Size	Frequency	F_{last}/F_i	Rank	Size	Frequency	F_{last}/F_i	Rank	Size	Frequency	F_{last}/F_i
1	6000	1	1	1	5444	1	1	1	8000	1	1	1	7000	1	1
2	4287	1	1	2	5000	1	1	2	4500	1	1	2	4861	1	1
3	2180	1	1	3	4500	1	1	3	4350	1	1	3	4500	1	1
4	2010	1	1	4	4000	1	1	4	4250	1	1	4	3500	1	1
5	1600	1	1	5	3000	1	1	5	4000	1	1	5	3300	1	1
6	1503	1	1	6	2900	1	1	6	3570	1	1	6	2800	1	1
7	1413	1	1	7	2400	1	1	7	2400	1	1	7	2600	1	1
8	1400	1	1	8	2341	1	1	8	2195	1	1	8	2500	1	1
9	1363	1	1	9	2146	1	1	9	2068	1	1	9	2450	1	1
10	1342	1	1	10	1849	1	1	10	2000	1	1	10	2400	1	1
	1851 Census				1861 Census				1871 Census				1881 Census		

Table 3: Summary statistics of the frequency and size firm distributions. (N = count of non-missing observations, mean = arithmetic mean, geo mean = geometric mean, har mean = harmonic mean, sum = sum of the values of non-missing observations, max = maximum, min = minimum, range = max − min, sd = standard deviation, variance = variance of each distribution, c = coefficient of variation (sd/mean), se(mean) = standard error of mean (sd/sqrt(n)), skewness = skewness of each distribution, kurtosis = kurtosis of each distribution, p1 = 1st percentile, p5 = 5th percentile, p10 = 10th percentile, p25 = 25th percentile, median = median (same as p50), p75 = 75th percentile, p90 = 90th percentile, p95 = 95th percentile, p99 = 99th percentile, iqr = interquartile range = p75 - p25). Notice that the sum of the frequency distributions is equal to the count (N) of the size distributions.

STATS	FREQUENCY				SIZE			
	1851	1861	1871	1881	1851	1861	1871	1881
N	390	488	468	576	181,900	175,451	138,105	181,125
mean	466.41	360.88	295.10	314.45	6.52	9.28	10.69	11.83
geo mean	4.85	5.12	5.58	5.16	3.09	3.89	4.12	4.10
har mean	1.89	1.98	2.11	2.01	2.14	2.54	2.60	2.57
sum	181,900	176,107	138,105	181,125	1,185,602	1,628,912	1,476,743	2,142,404
max	47,231	33,230	25,644	34,291	6,000	5,444	8,000	7,000
min	1	1	1	1	1	1	1	1
range	47,230	33,229	25,643	34,290	5,999	5,443	7,999	6,999
sd	3,480.76	2,602.95	2,021.41	2,413.43	32.81	45.85	56.69	64.38
variance	12,100,000	6,775,322	4,086,098	5,824,624	1,076.6	2,102.1	3,213.8	4,145.2
c	7.46	7.21	6.85	7.67	5.03	4.94	5.30	5.44
se(mean)	176.26	117.83	93.44	100.56	0.08	0.11	0.15	0.15
skewness	10.62	10.29	9.99	11.24	65.90	42.82	47.70	35.10
kurtosis	125.61	117.96	111.69	140.78	8,565.04	3,391.43	4,333.46	2,149.57
p1	1	1	1	1	1	1	1	1
p5	1	1	1	1	1	1	1	1
p10	1	1	1	1	1	1	1	1
p25	1	1	1	1	2	2	2	2
median	2	2	3	2	3	3	4	4
p75	10	11	14	12	6	7	8	8
p90	103	96	95	95	12	15	17	18
p95	584	498	469	434	19	25	29	32
p99	17960	11456	8990	9627	51	98	125	150
iqr	9	10	13	11	5	5	6	6

Table 4: Matrix Correlation of the frequency and size distributions.

CORR	Freq51	Freq61	Freq71	Freq81	Size51	Size61	Size71	Size81
Freq51	1							
Freq61	0.9858	1						
Freq71	0.9858	0.9997	1					
Freq81	0.9879	0.9998	0.9998	1				
Size51	0.838	0.8448	0.8522	0.8506	1			
Size61	0.8124	0.8566	0.8626	0.8592	0.9598	1		
Size71	0.81	0.8386	0.8477	0.8443	0.9734	0.9815	1	
Size81	0.7083	0.7548	0.7642	0.7595	0.9498	0.9782	0.9787	1

Table 5: Kolmogorov-Smirnov procedure to estimate the cut-off parameter u and maximum likelihood estimation of the shape parameter α using Clauset et al. MATLAB package [11], and Gillespie's R package [26].

	FREQUENCY			
	1851	**1861**	**1871**	**1881**
Clauset et al.				
$\hat{u}$ (frequency of firms)	3	1	1	1
$\hat{\alpha}$	1.433	1.469	1.45	1.467
Gillespie				
$\hat{u}$ (frequency of firms)	3	1	1	1
$\hat{\alpha}$	1.433	1.469	1.45	1.467
	SIZE			
	1851	**1861**	**1871**	**1881**
Clauset et al.				
$\hat{u}$ (size of firms)	10	10	10	10
$\hat{\alpha}$	2.481	2.321	2.242	2.144
Gillespie				
$\hat{u}$ (size of firms)	10	10	10	10
$\hat{\alpha}$	2.481	2.321	2.242	2.144

Table 6: Point estimation and 95% confidence intervals of Kolmogorov-Smirnov statistic to estimate the cut-off u parameter, and maximum likelihood estimation of the shape α parameter using Gillespie's R package [26] with tails or $Total\ obs - rank(\hat{u})$ and percentage of frequency and size of firms in the tail.

	FREQUENCY			
	1851	1861	1871	1881
Gillespie				
$\hat{u}$ (frequency of firms)	2.9 ($\approx$3)	1.32 ($\approx$1)	1.19 ($\approx$1)	1.18 ($\approx$1)
Confidence interval	(2.81-3.01)	(1.27-1.37)	(1.17-1.22)	(1.16-1.22)
Tail or $Total\ obs - rank(\hat{u})$	300	452	453	552
Confidence interval	(297-304)	(450-455)	(452-454)	(550-555)
% value in the tail (frequency of firms)	53.08	100	100	100
$\hat{\alpha}$	1.433	1.469	1.45	1.467
Confidence interval	(1.430-1.436)	(1.465-1.474)	(1.446-1.454)	(1.465-1.468)

	SIZE			
	1851	1861	1871	1881
Gillespie				
$\hat{u}$ (size of firms)	10.08 ($\approx$10)	9.97 ($\approx$10)	10.005 ($\approx$10)	9.79 ($\approx$10)
Confidence interval	(10.05-10.11)	(9.96-9.98)	(10.00-10.01)	(9.77-9.81)
Tail or $Total\ obs - rank(\hat{u})$	24,290	32,481	28,376	38,016
Confidence interval	(24,286-24,295)	(32,460-32,505)	(28,362-28,385)	(37,911-38,105)
% value in the tail (size of firms)	60.64	70.52	74.9	77.5
$\hat{\alpha}$	2.481	2.321	2.242	2.144
Confidence interval	(2.480-2.482)	(2.319-2.324)	(2.238-2.246)	(2.142-2.146)

non-parametric bootstrapping 95% confidence intervals in parentheses

Table 7: Loglikelihood ratio test for comparing the Pareto hypothesis with an exponential, a stretched exponential, and a Lognormal. The test results provided are the normalised ratio and the p-value calculated from it. A small p-value means the direction of change is caused not only by random variation. A large p-value means that the statistic cannot discern between each pair of distribution. Alstott package [2] in use with 5,000 simulations for each of the tails.

	FREQUENCY			
	1851	**1861**	**1871**	**1881**
EXPONENTIAL				
normalised ratio	9.482‡	10.27‡	10.795‡	9.935‡
p-value	0.00	0.00	0.00	0.00
STRETCHED EXPONENTIAL				
normalised ratio	1.715*	1.556	2.584‡	1.412
p-value	0.086	0.12	0.009	0.158
LOGNORMAL				
normalised ratio	-0.442	0.241	-0.786	0.158
p-value	0.658	0.809	0.432	0.874
	SIZE			
	1851	**1861**	**1871**	**1881**
EXPONENTIAL				
normalised ratio	20.309‡	32.801‡	29.6‡	38.74‡
p-value	0.00	0.00	0.00	0.00
STRETCHED EXPONENTIAL				
normalised ratio	16.612‡	24.21‡	20.813‡	20.788‡
p-value	0.00	0.00	0.00	0.00
LOGNORMAL				
normalised ratio	2.342†	10.141‡	1.469	0.808
p-value	0.019	0.00	0.142	0.419

*, $p < 0.10$; †, $p < 0.05$; ‡, $p < 0.01$

Table 8: Comparisons between 1881 (Own data) and 1987 (Robson and Gallagher data, [51]) tabulated in the size categories of Robson and Gallagher ([51]: Table 1). A typo for Total Firms 905,556 has been corrected, and a column for 1987 was calculated from total employment of 14,433,000 in Robson and Gallagher ([51]: Table 2). The difference between the UK and England and Wales using population estimates for the periods roughly equals 10%.

Firm size Bins	1881(England and Wales)			1987(UK)			Percentage Change 1987-1881		
	Firms	Employees	% Emp	Firms	Employees	% Emp	% Ch in Firms	% Ch in Emp	% Ch in % Emp
1-4	106,252	234,431	10.9	445,596	1,213,212	8.4	319.4	417.5	-23.2
5-9	37,922	247,681	11.6	234,079	1,559,844	10.8	517.3	529.8	-6.6
10-14	14,226	165,592	7.7	101,159	1,155,440	8.0	611.1	597.8	3.5
15-19	6,333	105,447	4.9	49,997	823,251	5.7	689.5	680.7	15.8
20-24	4,147	89,461	4.2	14,198	303,303	2.1	242.4	239.0	-49.7
25-29	2,191	58,467	2.7	8,636	231,088	1.6	294.2	295.2	-41.4
30-39	2,794	93,199	4.4	13,908	462,176	3.2	397.8	395.9	-26.4
40-49	1,521	65,777	3.1	7,780	332,189	2.3	411.5	405.0	-25.1
50-69	1,763	100,502	4.7	10,268	577,720	4.0	482.4	474.8	-14.7
70-99	1,100	89,482	4.2	7,068	563,277	3.9	542.5	529.5	-6.6
100-199	1,475	197,034	9.2	7,227	967,681	6.7	390.0	391.1	-27.1
200-299	545	126,632	5.9	2,014	476,619	3.3	269.5	276.4	-44.2
300-499	429	155,419	7.3	1,411	519,948	3.6	228.9	234.5	-50.4
500-699	190	106,860	5.0	456	259,974	1.8	140.0	143.3	-63.9
700-999	117	95,391	4.5	354	288,860	2.0	202.6	202.8	-55.1
1,000-1,499	66	76,559	3.6	270	274,417	1.9	309.1	258.4	-46.8
1,500-1,999	22	35,871	1.7	131	173,316	1.2	495.5	383.2	-28.3
2,000-4,999	29	81,599	3.8	301	592,163	4.1	937.9	625.7	7.6
5,000-9,999	3	17,000	0.8	123	491,062	3.4	4,000.0	2,788.6	328.5
+ 10,000	0	0	0.0	179	3,177,460	22.0	Inf	Inf	Inf
Total	181,125	2,142,404	100.0	905,155	14,443,000	100.0	399.7	574.1	0.0